Differential Privacy Working Group

Carol Hafford, NORC at the University of Chicago

May 27, 2021

Overview

Implementing Differential Privacy for the 2020 Census Data Products

Theoretical Understandings

Practical Implications for Hard-To-Count Populations

Implications for Redistricting

Implications for Race and Ethnic Total Populations

Stakeholder Engagement

Discussion

Implementing Differential Privacy for the 2020 Census Data Products

To begin, the NAC:

- Acknowledges the technical complexity of this effort and applauds the Census Bureau for its efforts
- Recognizes the Bureau for the improvements it has made in recent months, noting that the latest release of demonstration data has made significant improvements over the November 2020 Disclosure Avoidance System (DAS) release, in terms of the algorithm, many of the technical changes for small populations, and the increased Privacy Loss Budget
- Thanks the Bureau for sharing the results of its ongoing refinement of the DAS with the Differential Privacy Workgroup (DPWG), and seeking feedback from the data user community, tribal nations, civil-rights organizations, and others
- Appreciates the Bureau's efforts over the past year—amid the pandemic and a challenging legal environment—to conduct the decennial count and to fulfill constitutional requirement for congressional redistricting and the release of data products, including the PL94-171 (redistricting data)

Theoretical Understandings

The privacy interest is based on the Title 13 requirement not to publish exact identifying information

For the Census data products, differential privacy presents a trade-off:

- If the privacy loss parameter is set to favor utility, then the privacy benefits are lowered
- If the privacy loss parameter is set to favor privacy, then the utility benefits are lowered

The trade-off presents a quandary for:

- *Data utility and accuracy requirements*
- *Fitness-for-use (or quality of the data product) and protecting confidentiality*
- *What is ideal and what is practical*
- *Concerns with equality and equity*

Practical Implications

What data will change and what data won't?

What data won't change (remains invariant)?

- Total state population, which already is reflected in the apportionment data that the Census Bureau released on April 26, 2021
- Total number of housing units in each census block
- Number and type of each group quarters unit in each census block

What data will be subjected to the "noise" and may change (i.e., not be reported as enumerated)?

- Population
- Voting age population

What are the practical, and justice-oriented, implications of this trade-off for redistricting? And for hard-to-count population groups?

Redistricting Implications

Concerns raised in one series of analyses and case studies of redistricting use data (using April 2021 PPMFs and a conservative 4.5 PLB) found that:

- *As the largest minority population is prioritized in the privacy-loss budget, smaller minority groups often see more drastic fluctuation in population counts as a consequence*
  - In many cases, the largest minority population adds people at the expense of smaller minority populations
  - Loss of data – *people* – has implications for changes in federal funding and state spending, with downstream consequences for local and tribal governance
  - Decreased political representation
- *Diluting geographic compactness (which is essential to establishment of majority-minority voting districts)*
  - Reallocation of AI/AN populations at the census block level raises concerns about misrepresenting its geographic *compactness* (intuitively conceived as closeness to the shape and geographic center, but also a complex multidimensional concept) and diluting its statues as majority-minority district (if population moves across a county or state boundary)
- *Differential privacy may have significant negative consequences for redistricting less populated places and other less populated local government bodies, such as school boards and special districts*

Gall, M. (2021). *U.S. Census Bureau 2020 Disclosure Avoidance System and Implications for Redistricting.* Native American Rights Fund.

More Implications

Other priority uses of the PL94-171 data beyond redistricting are:

1. Federal, state and local appropriations
2. Data use by state, tribal, and local governments for evaluating public programs and making policy decisions based upon that data
3. General academic research

Implications for Race and Ethnic Total Populations

An iterative examination of the Census demonstration data products found that changes in total population affected the equal population requirement for redistricting

Analyses of the first public release of the Census Bureau's 2010 Demonstration Data Products in October 2019 and the fourth demonstration product released in November 2020 revealed differential, and concerning findings:

- *Analyses of the 1st demonstration product "showed small population counties tending to become more minority"*
- *But, analyses of the 4th demonstration product found that "small counties went from generally gaining Latino, non-Latino Asian, non-Latino Black, and non-Latino American Indian, Alaskan Native (AI/AN) population to losing racial and ethnic minority population on average"*
- *In other words, in the 4th demonstration product, researchers found more of a gain in non-Latino White population in these counties, meaning these counties are losing minority populations*

Source: MALDEF AAJC *Differential Privacy Preliminary Report* FINAL 4.5.2021.pdf (advancingjustice-aajc.org)

Implications for Race and Ethnic Total Populations, cont.

These concerns are supported by other analyses of the demonstration products that have raised questions about:

- Accuracy of counts of the American Indian/Alaska Native population on reservations and in smaller communities with populations less than 500 people (lower level geographies)
- Application of differential privacy to Black households and blocks that are already undercounted, especially in 2020
  - Black men between the ages of 18-49 have the highest omission rate of all racial and ethnic groups
  - Black children were undercounted by 6.3 percent in the 2010 census

Applying differential privacy to already significantly undercounted population groups—including American Indians and Alaska Natives, increases differentials in the count, as well as equal representation and voting rights under the law

Differential privacy has the potential of exacerbating unresolved issues related to the differential undercount for Black (including Black immigrant) populations

Sources:

Roubideaux, Y. & Evans-Lomayesva, G. (May 8, 2021). *What's the price of privacy? 2020 Census disclosure Avoidance System Impacts on Tribal Nation Census Data.* National Congress of American Indians;

National Urban League (communication).

Stakeholder Engagement and Communication

With these concerns, the NAC calls on the Bureau to increase it engagement with—not only with the scientific community–but with community members – as reflected across the NAC – and state, local, and tribal stakeholders who are most affected by these consequential shifts in counts and representation, potential funding and services that adherence to the differential privacy algorithm will create.

In closing, for the NAC to assist the Bureau in fulfilling the DPWG's 4th deliverable …

- to **develop strategies** for **communicating** the use of differential privacy for the 2020 Census data products
- and **provide recommendations** for **messaging and communication strategies** for **informing data users** (of all skill levels)

about what differential privacy is, the data protection it provides, the privacy/accuracy tradeoff,

we need to begin a frank and national discussion of its **strengths, limitations, and practical implications**.

Source: U.S. CENSUS BUREAU ADVISORY COMMITTEES. Implementing Differential Privacy for the 2020 Census Data Products. National Advisory Committee Census Scientific Advisory Committee. Approved on February 27, 2020.